CMPUT 466/566: Machine Learning, Winter 2024
Tutorial 3

Shuai Liu

1 Multivariate Calculus

Definition 1 (partial derivative)**.** *For a function $f : \mathbb{R}^d \to \mathbb{R}$, and standard Euclidean basis $\{e_i\}_{i=1}^d$, if the following limit of function exists at a, we say f has partial derivative w.r.t. x_i at a*

$$\lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}.$$

We denote the above limit by $\frac{\partial f(a)}{\partial x_i}$ or $\frac{\partial}{\partial x_i} f(a)$.

Remark 1. In this course, it is ok to treat everything else as a constant and only take derivative w.r.t. the target variable.

Example 1. *Let $a \in \mathbb{R}^d$ and $f(x) = a^\top x$. Then*

$$\frac{\partial}{\partial x_i} f(x) = \frac{\partial}{\partial} (\sum_{j=1}^d a_j x_j) = a_i.$$

Definition 2 (gradient)**.** *For a function $f : \mathbb{R}^d \to \mathbb{R}$, assume the partial derivatives of f all exist in a neighborhood of x and are also continuous at x. Then f is differentiable and its gradient satisfies*

$$\nabla_x f(x) = \left(\frac{\partial f}{\partial x} \right)^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}$$

Example 2. *Let $a \in \mathbb{R}^d$ and $f(x) = a^\top x$. Then* $\nabla_x f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^d a_i x_i \right) \\ \frac{\partial f}{\partial x_2} \left(\sum_{i=1}^d a_i x_i \right) \\ \vdots \\ \frac{\partial f}{\partial x_d} \left(\sum_{i=1}^d a_i x_i \right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}$

Definition 3 (Jacobian)**.** *For a differentiable function $f \in \mathbb{R}^n \to \mathbb{R}^m$, let the output of function be $(f_1(x), ..., f_m(x))^\top$ the Jacobian of the function is defined as*

$$J_f(x) = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \vdots \\ \frac{\partial f_m}{\partial x} \end{bmatrix}$$

Proposition 1 (Chain rule)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ are both differentiable with Jacobians $J_g(f(x))$ and $J_f(x)$. Then the derivative*

$$\frac{\partial g \circ f}{\partial x} = \sum_{j=1}^m \frac{\partial g(f_1(x), ..., f_m(x))}{\partial f_j(x)} \frac{\partial f_j(x)}{\partial x}.$$

Proposition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ are both differentiable with Jacobians $J_g(f(x))$ and $J_f(x)$. Then $g \circ f$ has Jacobian $J_{g\circ f}(x) = J_g(f(x))J_f(x)$.*

Proof. By definition of function composition, $g \circ f$ is a function that maps from $\mathbb{R}^n \to \mathbb{R}^p$. As usual, we denote the output of $g \circ f$ as $((g \circ f)_1(x), ..., (g \circ f)_p(x))^\top$. Then the j-th element in the i-th row of the Jacobian of $J_{g\circ f}$ is that

$$\begin{aligned}
[J_{g\circ f}(x)]_{ik} &= \frac{\partial (g\circ f)_i}{\partial x_k} \\
&= \frac{\partial g_i(f(x))}{\partial x_k} \\
&= \frac{\partial g_i(f_1(x), ..., f_m(x))}{\partial x_k} \\
&= \sum_{j=1}^{m} \frac{\partial g_i(f_1(x), ..., f_m(x))}{\partial f_j(x)} \frac{\partial f_j(x)}{\partial x_k} && \text{(Chain rule)} \\
&= \sum_{j=1}^{m} [J_g(f(x))]_{ij}[J_f(x)]_{jk} \\
&= [J_g(f(x))J_f(x)]_{ik}
\end{aligned}$$

□

Definition 4 (Hessian)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice differentiable, the Hessian is defined to be*

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}.$$

If f is twice continuously differentiable, then $\nabla^2 f(x)$ is symmetric.

2 Convex Analysis

Definition 5 (Convex set)**.** *A set $V \subseteq \mathbb{R}^d$ is convex if for all $x, y \in V$ and $\gamma \in (0,1)$, we have that $\gamma x + (1-\gamma)y \in V$.*

Remark 2. See Wikipedia on convex sets for examples of convex sets.

Definition 6 (Convex function)**.** *A function $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$ is convex if V is a convex set and for all $x, y \in V$, $f(\gamma x + (1-\gamma)y) \leq \gamma f(x) + (1-\gamma) f(y)$.*

> For definitions of neighborhood, open set, interior and some simple examples, please refer to the appendix. A good reference is Page 32 of Baby Rudin.

Proposition 3 (First order condition)**.** *Assume $f : \mathcal{V} \subset \mathbb{R}^d \to \mathbb{R}$ is differentiable in the interior of V, which we denote by V°, then f is convex if and only if for all $x \in V, y \in V^\circ$*

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x)$$

Remark 3. The reason why we introduce the interior is because f cannot be differentiable on the boundary. Think about if a function $g : [a, b] \to \mathbb{R}$ can be differentiable on the endpoint (a or b).

Proposition 4 (Second order condition)**.** *Assume $f : V \subset \mathbb{R}^d \to \mathbb{R}$ is twice-continuously-differentiable in the interior of V, which we denote by V°, then f is convex if and only if for all $x \in V^\circ$*

$$\nabla^2 f(x) \succeq 0$$

Definition 7 (Local minima)**.** *For a function $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$, $x^* \in V$ is a local minima of f if there exists a neighborhood $\mathcal{U} \subseteq V$ of x^* such that for all $y \in \mathcal{U}$, $f(y) \geq f(x^*)$.*

Definition 8 (Global minima)**.** *For a function $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$, $x^* \in V$ is a global minima of f if for all $y \in V$, $f(y) \geq f(x^*)$.*

Proposition 5. *For a convex function $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$, any local minima of f is a global minima.*

Proposition 6 (Condition 1 for minima)**.** *Let $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$ be a function that is differentiable on V°. If x^* is a local minima of f then*

$$\nabla f(x^*) = 0.$$

Furthermore, if f is convex, then x^* is a global minima of f if and only if

$$\nabla f(x^*) = 0.$$

Proposition 7 (Condition 2 for minima)**.** *Let $f : V \subseteq \mathbb{R}^d \to \mathbb{R}$ be a function that is twice-continuously-differentiable on V° and $\nabla^2 f(x) \succeq 0$ for all x in a neighborhood $\mathcal{U} \subseteq V$ of $x^* \in V$. Then x^* is the local minima of f if and only if*

$$\nabla f(x^*) = 0.$$

3 Appendix

Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and consider the space $(\mathbb{R}^d, \|\cdot\|)$.

Definition 9 (Neighborhood)**.** *For a point $v \in \mathbb{R}^d$, $\mathcal{U}$ is called a neighborhood of v if $v \in \mathcal{U}$ and there exists $r > 0$ such that the open ball centered at v with radius being r subsets $\mathcal{U}$.*

$$B(r) := \{x : \|x - v\| < r\} \subseteq \mathcal{U}$$

Remark 4. For example, an open ball centered at v is a neighborhood of v. This is a generalization of open interval in $\mathbb{R}$. Think about $\mathbb{R}$, the neighborhood of $x \in \mathbb{R}$ is an open interval that contains $\mathbb{R}$.

Definition 10 (Open set)**.** *A set $\mathcal{V} \subseteq \mathbb{R}^d$ is an open set if for all $v \in \mathcal{V}$, there exists a neighborhood $\mathcal{U}_v$ of v such that $\mathcal{U}_v \subseteq \mathcal{V}$.*

Remark 5. An open ball centered at v is an open set. Think about $\mathbb{R}$, an union of open intervals, $\cup_{i=1}^{\infty} \mathcal{I}_i$ where $\mathcal{I}_i$ are open intervals, is an open set. In fact, all open sets in $\mathbb{R}$ can be written as the union of open intervals.

Definition 11 (Interior)**.** *For a set $\mathcal{V} \subseteq \mathbb{R}^d$, the interior $\mathcal{V}^\circ$ is the set of points that have a neighborhood that subsets $\mathcal{V}$, that is,*

$$\mathcal{V}^\circ = \{x \in \mathcal{V} : \exists\, \mathcal{U}_x \text{ neighborhood of } x \text{ s.t. } \mathcal{U}_x \subseteq \mathcal{V}\}.$$

Remark 6. Picture whatever closed shape in your mind, for example, a heart. Then exclude the boundaries, every point inside forms the interior of that shape. For a closed ball,

$$K(r) := \{x : \|x - v\| \leq r\},$$

the interior is $B(r)$. Think about $\mathbb{R}$, the interior of a closed interval is just the open interval version of it (which is the largest open interval subsets it) and the interior of $\mathbb{N}$ is empty.

Definition 12 (Boundary)**.** *For a set $\mathcal{V} \subseteq \mathbb{R}^d$, the boundary of $\mathcal{V}$, which we denote by $\partial\mathcal{V}$, is $\partial\mathcal{V} := \mathcal{V} \backslash \mathcal{V}^\circ$.*